

AD-A056 722

WISCONSIN UNIV-MADISON DEPT OF INDUSTRIAL ENGINEERING

F/G 12/1

STATISTICAL ANALYSIS OF THE OUTPUT DATA FROM TERMINATING SIMULA--ETC(U)

APR 78 A M LAW

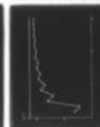
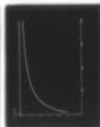
N00014-76-C-0403

UNCLASSIFIED

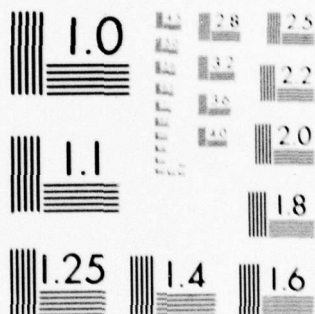
TR-78-4

NL

1 OF 1
AD
A056 722



END
DATE
FILMED
9-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

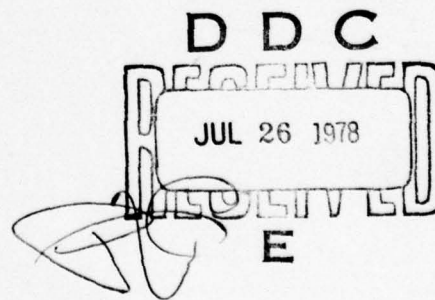
2 (12)

DEPARTMENT OF INDUSTRIAL ENGINEERING

AD A 056722



AD No. _____
DDC FILE COPY.



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

 **UNIVERSITY OF
WISCONSIN-MADISON**

78 07 25 002

12

AD A 056722

AD No. /
DDC FILE COPY

6 STATISTICAL ANALYSIS OF THE OUTPUT DATA
FROM TERMINATING SIMULATIONS.

10 by
Averill M. Law

11 Apr 1978

78-4

9 Technical rept.

14 TR-78-4

15 N00014-76-C-0403

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DOC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

12 34 p.

DDC
JUL 26 1978
E

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

This research was supported by the Office of Naval Research under contract N00014-76-C-0403 (NR 047-145).

Reproduction in whole or in part is permitted for any purpose of the United States Government.

408 848
78 07 25 002

STATISTICAL ANALYSIS OF THE OUTPUT DATA
FROM TERMINATING SIMULATIONS*

Averill M. Law
University of Wisconsin

ABSTRACT

In this paper we precisely define the two types of simulations (terminating and steady-state) with regard to analysis of simulation output and discuss some common measures of performance for each type. In addition, we conclude from talking to a large number of simulation practitioners, that a significant proportion of simulations in the real world are of the terminating type. This is contrary to the impression one gets from reading the simulation literature, where the steady-state case is almost exclusively considered.

Although analyses of terminating simulations are considerably easier than are those of steady-state simulations, they have not received a careful treatment in the literature. We discuss and give empirical results for fixed sample size, relative width, and absolute width procedures which can be used for constructing confidence intervals for measures of performance in the terminating case.

*This research was supported by the Office of Naval Research under contract N00014-76-C-0403 (NR 047-145).

1. Types of Simulations with Regard to Analysis of the Output

We begin by giving a precise definition of the two types of simulations with regard to analysis of the output data (cf.

Gafarian and Ancker [4] and Kleijnen [6]). A terminating simulation is one for which any quantities of interest are defined relative to the interval of simulated time $[0, T_E]$, where T_E , a possibly degenerate random variable (r.v.), is the time that a specified event E occurs. The following are some examples of terminating simulations:

- a) Consider a retail establishment (e.g., a bank) which closes each evening (physically terminating). If the establishment is open from 9 to 5, then the objective of a simulation might be to estimate the quality of customer service over the 8 hour period. In this case, $E = \{8 \text{ hours of simulated time have elapsed}\}$.
- b) Consider a telephone exchange which is always open (physically nonterminating). Since the arrival rate of calls changes with the time of day, day of the week, etc., it is unlikely that a steady-state measure of performance (see Section 2), which is defined as a limit as time goes to infinity, will exist. In this case the objective of a simulation might be to study the system during the period of peak loading, say, of length t hours, and $E = \{t \text{ hours of simulated time have elapsed}\}$.
- c) A system consists of mechanical and electronic components, each of which is subject to failure. The system itself fails

when certain specified subsets of the components fail and the objective of a simulation might be to estimate some characteristic of the time to failure of the whole system. In this case, $E = \{\text{the system fails}\}$.

A steady-state simulation is one for which the quantity of interest is defined as a limit as the length of the simulation goes to infinity. Since there is no natural event to terminate the simulation, the length of the simulation is made large enough to get a "good" estimate of the quantity of interest. Alternatively, the length of the simulation could be determined by cost considerations. The following are some examples of steady-state simulations:

- a) Consider a computer manufacturer who constructs a simulation model of a proposed computer system. Rather than use data from the arrival process of an existing computer system as input to the model, he typically assumes that jobs arrive in accordance with a Poisson process with rate equal to the predicted arrival rate of jobs during the period of peak loading. He is interested in estimating the response time of a job after the system has been running long enough so that initial conditions (e.g., the number of jobs in the system at time 0) no longer have any effect. (Assuming that the arrival rate is constant over time allows steady-state measures to exist.)
- b) A chemical manufacturer constructs a simulation model of a proposed chemical process operation. The process, when in

operation, will be subject to randomly occurring breakdowns. The input rate of raw materials to the process and the controllable parameters of the process are both assumed to be stationary with respect to time. The company would like to estimate the production rate after the process has been running long enough so that initial conditions no longer have any effect.

The remainder of this paper is organized as follows. In Section 2 we discuss some common measures of performance for both types of simulations and in Section 3 we present our findings on the relative occurrence of each type in the real world. A number of procedures which can be used to construct confidence intervals for terminating simulations are discussed in Section 4 and, finally, in Section 5 we summarize our findings.

2. Measures of System Performance

To the best of our knowledge, nowhere in the simulation literature are measures of performance for terminating simulations explicitly defined. In this section we define and contrast several common measures of performance for terminating and steady-state simulations by means of examples. (Because of the diversity of terminating simulations, it is not possible to give one definition that fits all cases.) For the examples that we consider, it is possible to compute analytically measures of performance. Furthermore, the same examples will be considered in Section 4 where we discuss stopping rules for terminating simulations.

A. Averages

Consider first the stochastic process $\{D_i, i \geq 1\}$ for the M/M/1 queue with $\rho < 1$, where D_i is the delay in queue of the i th customer. The objective of a terminating simulation of the M/M/1 queue might be to estimate the expected average delay of the first m customers given that the number of customers in the system at time 0, $N(0)$, is zero. The desired quantity, which we denote by $d(m|N(0)=0)$, is then given by

$$\begin{aligned} d(m|N(0)=0) &= E\left[\sum_{i=1}^m D_i/m | N(0)=0\right] \\ &= \sum_{i=1}^m E[D_i | N(0)=0]/m . \end{aligned}$$

(Alternatively, if one is interested in estimating the expected average delay of all customers who arrive and are served in the time interval $[0, t]$, then the desired quantity is given by

$$d(t|N(0)=0) = E\left[\sum_{i=1}^{M(t)} D_i/M(t) | N(0)=0\right] ,$$

where $M(t)$ (a r.v.) is the number of customers who arrive and are served in the interval $[0, t]$. Note that in this case the expectation and summation are not interchangeable. Thus, the label "expected average delay" is more general than "average expected delay".)

Note also that the quantity $d(m|N(0)=0)$, which is often called a transient characteristic of the stochastic process $\{D_i, i \geq 1\}$, explicitly depends on the state of the system at time 0; i.e., $d(m|N(0)=i) \neq d(m|N(0)=j)$ for $i \neq j$.

The objective of a steady-state simulation of $\{D_i, i \geq 1\}$ for the M/M/1 queue would be to estimate the steady-state expected average delay d , which is given by

$$d = \lim_{m \rightarrow \infty} d(m|N(0)=i) \quad \text{for any } i=0,1,\dots$$

Observe that d is independent of $N(0)$. In Figure 1 we plot $d(m|N(0)=0)$ as a function of m . (The arrival rate $\lambda = 1$ and the service rate $\mu = 10/9$, so $\rho = .9$.) The horizontal line that $d(m|N(0)=0)$ asymptotically approaches is at height d .

As a second example consider the stochastic process $\{E_i, i \geq 1\}$ for an (s,S) inventory system with zero delivery lag and backlogging, where E_i is the expenditure in the i th period. This system is described in detail in Law [8]. A possible objective of a terminating simulation would be to estimate the expected average cost for the first m periods given that the inventory level at the beginning of period 1, I_1 , is S :

$$e(m|I_1=S) = E \left[\sum_{i=1}^m E_i / m | I_1=S \right]$$

The objective of a steady-state simulation of $\{E_i, i \geq 1\}$ would be to estimate the steady-state expected average cost:

$$e = \lim_{m \rightarrow \infty} e(m|I_1=i) \quad \text{for any } i=0,\pm 1,\pm 2,\dots$$

We plot $e(m|I_1=S)$ as a function of m and also e in Figure 2.

Our third example is quite different from the first two. In the reliability model shown in Figure 3 it is desired to send a message from A to B; this will occur if component 1 works and either component 2 or 3 works. If T is the time to failure of the whole system and T_i the time to failure of component i ($i=1,2,3$), then

$$T = \min[T_1, \max(T_2, T_3)] .$$

We further assume that the T_i 's are independent r.v.'s and each T_i has a Weibull distribution with shape parameter $\alpha=.5$ and scale parameter $\beta=1$. (A distributional assumption for the T_i 's is needed in Section 4 where we present simulation results for this model.) The objective of a terminating simulation might be to estimate the expected time to failure of the system given that all components are new, $E(T|\text{all components are new})$. If we assume that the system is not repaired when it fails, then a steady-state simulation makes no sense for this system. Such could be the case if this system were part of a space probe.

B. Proportions

The usual criterion for comparing two or more systems is some sort of average behavior. However, different kinds of information may be of more value in some situations. For example, a bank manager may be concerned with estimating the proportion of customers who experience a delay in excess of 5 minutes. Since proportions are really just a special case of averages, we will illustrate them by means of the M/M/1 example.

In a terminating simulation of the M/M/1 queue the objective might be to estimate, instead of an expected average delay, the expected proportion of the first m customers whose delay is less than or equal to x (a specified number) given that $N(0)=0$. Denote the desired quantity by $P(m,x|N(0)=0)$ and let

$$Y_i(x) = \begin{cases} 1 & \text{if } D_i \leq x \\ 0 & \text{if } D_i > x \end{cases} \quad \text{for } i=1,2,\dots$$

Then $P(m,x|N(0)=0)$ is given by

$$\begin{aligned} P(m,x|N(0)=0) &= E \left[\sum_{i=1}^m Y_i(x) / m \mid N(0)=0 \right] \\ &= \sum_{i=1}^m E[Y_i(x) | N(0)=0] / m \\ &= \sum_{i=1}^m P\{D_i \leq x | N(0)=0\} / m . \end{aligned}$$

For a steady-state simulation, the objective would be to estimate the steady-state expected proportion of customers whose delay is less than or equal to x :

$$P(x) = \lim_{m \rightarrow \infty} P(m,x|N(0)=i) \quad \text{for any } i=0,1,\dots$$

3. Relative Importance of Each Type of Simulation

Reading the simulation literature leads one to think that steady-state simulations are more important; almost every paper written on the analysis of simulation output data deals with the

steady-state case. This may be a carry-over from mathematical queueing theory where only a steady-state analysis is generally possible. However, we have discovered by talking to a large number of simulation practitioners that a significant proportion of simulations in the real world are actually of the terminating type. The following are some reasons why a steady-state analysis may not be appropriate:

- a) The system under consideration is physically terminating. In this case, letting the length of a simulation be arbitrarily large makes no sense.
- b) The input distributions for the system change over time. In this case, steady-state measures of performance will probably not exist.
- c) One is often interested in studying the transient behavior of a system even if steady-state measures of performance exist.

4. Stopping Rules for Terminating Simulations

In the following three subsections we consider procedures which can be used to construct confidence intervals (c.i.'s) for measures of performance for terminating simulations. We will not consider the steady-state case since it has been widely discussed in the simulation literature. For surveys of fixed sample size and sequential procedures which can be used to construct c.i.'s for steady-state measures of performance, see Law [9] and Law and Kelton [10], respectively. The random numbers used in the remainder of this paper were generated from the generator discussed in [8].

A. Fixed Sample Size Procedures

Suppose we make n independent replications of a terminating simulation. The independence among replications is accomplished by using different random numbers for each replication and by starting each one with the same initial conditions. If X_i is the estimator of interest from the i th replication ($i=1,2,\dots,n$), then the X_i 's are independent identically distributed (i.i.d.) r.v.'s. (For the M/M/1 queue, X_i might be the average $\sum_{j=1}^m D_j/m$ or the proportion $\sum_{j=1}^m Y_j(x)/m$.) If, in addition, the X_i 's are normally distributed, then a $100(1-\alpha)\%$ ($0 < \alpha < 1$) c.i. for $\mu = E(X)$ is given by

$$\bar{X}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{s^2(n)/n}, \quad (1)$$

where $\bar{X}(n)$ and $s^2(n)$ are the usual sample mean and variance, respectively, and $t_{n-1, 1-\alpha/2}$ is the $1 - \alpha/2$ point for a t distribution with $n-1$ degrees of freedom.

In practice the X_i 's will not be normally distributed and the c.i. given by (1) will be only approximate. To investigate the effect of nonnormality, we simulated the three stochastic models of Section 2. For the M/M/1 queue with $\rho = .9$, the (s,S) inventory system, and the reliability model, respectively, the quantities of interest were $d(25|N(0)=0) = 2.12$, $e(12|I_1=S) = 99.52$, and $E(T|\text{all components are new}) = .778$. (See [8] for a discussion of how to compute the first two quantities.) For each model we performed 500 independent simulation experiments, for each experiment we considered $n = 5, 10, 20, 40$, and for each n we used (1)

to construct a 90% c.i. for the desired quantity. In Tables 1,2, and 3 we give the proportion, \hat{p} , of the 500 c.i.'s which covered the desired quantity, a 90% c.i. for the true coverage, and the average value of the c.i. half length divided by the point estimate over the 500 experiments for the three models. The 90% c.i. for the true coverage was computed from

$$\hat{p} \pm 1.645\sqrt{\hat{p}(1-\hat{p})/500} .$$

Observe that for the M/M/1 queue and the (s,S) inventory system the coverages are quite close to 90%, but for the reliability model there is a significant degradation in coverage, apparently caused by a severe departure from normality. To see if this is indeed the case, we generated 1000 X_i 's for each stochastic model and estimated the skewness and kurtosis. These estimates, which are presented in Table 4, indicate that the X_i 's for the reliability model are considerably more nonnormal than are those for the other two models. This conclusion was reinforced by plotting histograms for the three sets of data.

B. Relative Width Procedures

One disadvantage of the fixed sample size approach to constructing a c.i. is that the simulator has no control over the c.i. half length; for fixed n , the half length will depend on the population variance $\sigma^2 = \text{Var}(X)$. In this subsection we consider two sequential procedures which allow one to specify the "relative precision" of

a c.i. Both assume that X_1, X_2, \dots is a sequence of i.i.d. r.v.'s which need not be normal.

The first procedure has been suggested for use in several different contexts; see Iglehart [5], Lavenberg and Sauer [7], and Thomas [13]. The objective of the procedure is to construct a $100(1-\alpha)\%$ c.i. for μ such that the difference between the point estimator $\bar{X}(n)$ and μ is no more than $100 \gamma\%$ of $\bar{X}(n)$, that is,

$$|\bar{X}(n) - \mu| \leq \gamma |\bar{X}(n)| \quad \text{for } 0 < \gamma < 1. \quad (2)$$

Choose an initial sample size $n_0 \geq 2$, let

$$\delta_{r,1}(n, \alpha) = t_{n-1, 1-\alpha/2} \sqrt{s^2(n)/n},$$

and let

$$N_{r,1}(\gamma, \alpha) = \min \left\{ n: n \geq n_0, s^2(n) > 0, \frac{\delta_{r,1}(n, \alpha)}{|\bar{X}(n)|} \leq \gamma \right\}. \quad (3)$$

(Note that $N_{r,1}(\gamma, \alpha)$, which is the required number of replications, is a r.v.) Then use

$$I_{r,1}(\gamma, \alpha) = [\bar{X}(N_{r,1}(\gamma, \alpha)) - \delta_{r,1}(N_{r,1}(\gamma, \alpha), \alpha), \bar{X}(N_{r,1}(\gamma, \alpha)) + \delta_{r,1}(N_{r,1}(\gamma, \alpha), \alpha)] \quad (4)$$

as an approximate $100(1-\alpha)\%$ c.i. for μ . It easily follows from

(3) and (4) that $I_{r,1}(\gamma, \alpha)$ satisfies the criterion given by (2).

Furthermore, using an argument similar to the one employed by

Lavenberg and Sauer in the context of the regenerative method for

steady-state simulations, we have been able to prove the following

theorem.

Theorem 1. If $\mu \neq 0$ and $0 < \sigma^2 < \infty$, then $\lim_{\gamma \rightarrow 0^+} P\{\mu \in I_{n,1}(\gamma, \alpha)\} = 1 - \alpha$.

The objective of the second procedure, which is due to Nadas [11], is to construct a c.i. such that

$$|\bar{X}(n) - \mu| \leq \gamma |\mu| \quad \text{for } 0 < \gamma < 1. \quad (5)$$

Let

$$v^2(n) = \left\{ 1 + \sum_{i=1}^n [X_i - \bar{X}(n)]^2 \right\} / n = (1/n) + (n-1)s^2(n)/n,$$

$$\delta_{n,2}(n, \alpha) = t_{n-1, 1-\alpha/2} \sqrt{v^2(n)/n},$$

and

$$N_{n,2}(\gamma, \alpha) = \min \left\{ n : n \geq n_0, \frac{\delta_{n,2}(n, \alpha)}{|\bar{X}(n)|} \leq \gamma \right\}.$$

Then use

$$I_{n,2}(\gamma, \alpha) = \left[\frac{\bar{X}(N_{n,2}(\gamma, \alpha))}{1 + \gamma}, \frac{\bar{X}(N_{n,2}(\gamma, \alpha))}{1 - \gamma} \right] \quad (6)$$

as an approximate $100(1-\alpha)\%$ c.i. for μ . From (6) it is easy to show that $I_{n,2}(\gamma, \alpha)$ satisfies the criterion given by (5). Furthermore, the following theorem was proved by Nadas.

Theorem 2. If $\mu \neq 0$ and $0 < \sigma^2 < \infty$, then $\lim_{\gamma \rightarrow 0^+} P\{\mu \in I_{n,2}(\gamma, \alpha)\} = 1 - \alpha$.

In order to compare the two procedures and to determine the effect of non-infinitesimal γ on coverage, we once again simulated the three stochastic models. For each model we performed 500 independent experiments, for the M/M/1 queue and the reliability model

we considered $\gamma = .2, .1, .05$ for each experiment, and for the inventory system we considered $\gamma = .2, .1, .05, .025, .0125, .00625$ for each experiment. In all cases, $n_0 = 5$. In Tables 5, 6, and 7 we give point estimates and 90% c.i.'s for the true coverages, point estimates and 90% c.i.'s for $E\{N_{r,i}(\gamma, \alpha)\} (i=1,2)$, and the average c.i. half lengths over the 500 experiments. We considered more values of γ for the inventory system because it appeared from our empirical results that a smaller γ is required for the coverage ultimately to converge to the desired level. (A smaller γ is required for this model to get a large value of $N_{r,i}(\gamma, \alpha)$.) Note also that convergence of coverage does not appear to be monotone.

We repeated the above 500 experiments using the same random numbers and $n_0 = 2$. For procedure 2 the results were identical; however, for procedure 1 there was a significant degradation in coverage due to premature stopping on replications 2,3, or 4. For example, the coverage for the M/M/1 queue with $\gamma = .2$ was .798.

c. Absolute Width Procedures

In this subsection we present two procedures which allow one to construct a $100(1-\alpha)\%$ c.i. for μ such that

$$|\bar{X}(n) - \mu| \leq \sigma, \quad (7)$$

where σ is a specified positive number.

The first procedure, which is due to Chow and Robbins [1], assumes that X_1, X_2, \dots is a sequence of i.i.d. r.v.'s. Choose $n_0 \geq 2$.

Let $v^2(n)$ be defined as in Subsection 4.B, let

$$N_{\alpha,1}(c,\alpha) = \min \left\{ n: n \geq n_0, v^2(n) \leq \frac{c^2 n}{(t_{n-1,1-\alpha/2})^2} \right\},$$

and then use

$$I_{\alpha,1}(c,\alpha) = [\bar{X}(N_{\alpha,1}(c,\alpha)) - c, \bar{X}(N_{\alpha,1}(c,\alpha)) + c]$$

as an approximate $100(1-\alpha)\%$ c.i. for μ . It is clear that $I_{\alpha,1}(c,\alpha)$ satisfies the criterion given by (7). The following theorem was proved by Chow and Robbins.

Theorem 3. If $0 < \sigma^2 < \infty$, then $\lim_{c \rightarrow 0^+} P\{\mu \in I_{\alpha,1}(c,\alpha)\} = 1 - \alpha$.

For an empirical evaluation of the above procedure under the assumption that the X_i 's are normal, see Starr [12].

The second procedure, which is due to Dudewicz [2], assumes that the X_i 's are i.i.d. normal r.v.'s. Initially make $n_0 (n_0 \geq 2)$ replications of the simulation and compute $\bar{X}(n_0)$ and $s^2(n_0)$. Let

$$N_{\alpha,2}(c,\alpha) = \max\{n_0+1, \lceil w^2 s^2(n_0) \rceil\},$$

where $w = t_{n_0-1,1-\alpha/2}/c$ and $\lceil z \rceil$ is the smallest integer $\geq z$. Make $N_{\alpha,2}(c,\alpha) - n_0$ additional replications of the simulation, let

$$\bar{Y}(N_{\alpha,2}(c,\alpha) - n_0) = \frac{1}{N_{\alpha,2}(c,\alpha) - n_0} \sum_{i=n_0+1}^{N_{\alpha,2}(c,\alpha)} X_i,$$

and let $\tilde{\bar{X}}(N_{\alpha,2}(c,\alpha)) = a_1 \bar{X}(n_0) + a_2 \bar{Y}(N_{\alpha,2}(c,\alpha) - n_0)$, where

$$a_1 = \frac{n_0}{N_{\alpha,2}(c,\alpha)} \left[1 + \sqrt{1 - \frac{N_{\alpha,2}(c,\alpha)}{n_0} \left(1 - \frac{N_{\alpha,2}(c,\alpha) - n_0}{w^2 s^2(n_0)} \right)} \right]$$

and $\alpha_2 = 1 - \alpha_1$. Then use

$$I_{\alpha,2}(c,\alpha) = [\bar{X}(N_{\alpha,2}(c,\alpha)) - c, \bar{X}(N_{\alpha,2}(c,\alpha)) + c]$$

as an approximate $100(1-\alpha)\%$ c.i. for μ . Dudewicz has proved the following theorem.

Theorem 4. $P\{\mu \in I_{\alpha,2}(c,\alpha)\} = 1 - \alpha$ for all $c > 0$.

To compare the sequential procedure of Chow and Robbins and the two-stage procedure of Dudewicz, we performed 500 independent experiments for each model. To make the absolute width results somewhat comparable to the relative width results, we chose the values of c to correspond to the values of γ ; that is, for each γ we chose $c = \gamma\mu$. For the Chow and Robbins procedure we chose $n_0 = 5$ and for the Dudewicz procedure we considered $n_0 = 15, 30$, and 60 . (Dudewicz [3] recommended that n_0 be at least 12.) The results of the simulation experiments for the three models are given in Tables 8, 9, and 10.

5. Summary and Conclusions

We have defined terminating and steady-state simulations and have discussed some common measures of performance for each type. In addition, we have concluded from talking with simulation practitioners that a significant proportion of real-world simulations are of the terminating type. This is fortunate because it means that classical statistical analysis for i.i.d. observations (e.g., confidence intervals, hypothesis testing, ranking and selection, etc.)

is applicable to analyzing many simulations. On the other hand, in the steady-state case there is still not a totally acceptable procedure even for the relatively simple problem of constructing a c.i. for a steady-state expected average.

We have also considered procedures for constructing c.i.'s for terminating simulations. If one is performing an exploratory experiment where precision of the c.i. may not be overwhelmingly important, then we recommend using a fixed sample size procedure. However, if the X_i 's are highly nonnormal and if the number of replications n is too small, then the actual coverage of the constructed c.i. may be considerably lower than that desired (see Table 3).

If one wants a c.i. having half length that is small relative to the point estimate, then a relative width procedure may be used. We recommend using Procedure 2 (due to Nadas) with $n_0 \geq 5$. Procedure 2 appears to give slightly better coverage, its criterion (see (5)) is more intuitive than the criterion of Procedure 1 (see (2)), and Procedure 2 does not seem subject to premature stopping even for $n_0 = 2$. (On the other hand, Procedure 1 uses a more intuitive expression to construct a c.i.)

If one wants a c.i. for which the half length is a specified number, then an absolute width procedure may be used. We recommend using the Chow and Robbins procedure with $n_0 \geq 5$. Their procedure generally requires a smaller average sample size, the variance of

the sample size is smaller, and its coverage seems to be less affected by departures from normality (see Table 10).

In general, we believe that relative width procedures are more useful than absolute width procedures due to the difficulty in specifying the absolute width σ for most simulation experiments. When using either the Nadas procedure or the Chow and Robbins procedure, we believe that it is advisable to choose a γ or σ which will cause the procedure to run until the sample is at least of moderate size; perhaps, at least 30. (Since both procedures are based on the central limit theorem, it is unreasonable to think that they will work well in general for a small sample size; see the results for $\gamma = .025$ in Table 6.) Finally, we mention that precise c.i.'s may be unaffordable in the real world due to the high cost of making a single replication.

Acknowledgements

The author would sincerely like to thank David Kelton for his programming assistance and for his reading of the manuscript. The author is also grateful to Stephen Lavenberg (IBM Watson Research Center), Richard Smith (IBM Boulder), and Charles White (Dupont) for their discussions of simulation at their companies.

References

1. Chow, Y.S. and H. Robbins, "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean," Ann. Math. Statist. 36, 457-462 (1965).
2. Dudewicz, E.J., "Statistical Inference with Unknown and Unequal Variances," Transactions of the Annual Quality Control Conference of the Rochester Society for Quality Control 28, 71-85 (1972).
3. Dudewicz, E.J., Personal Communication (1977).
4. Gafarian, A.V. and C.J. Ancker, Jr., "Mean Value Estimation from Digital Computer Simulation," Operations Res. 14, 25-44 (1966).
5. Iglehart, D.L., "The Regenerative Method for Simulation Analysis," Technical Report No. 86-20, Control Analysis Corporation, Palo Alto, California (1975).
6. Kleijnen, J.P.C., "The Statistical Design and Analysis of Digital Simulation: A Survey," Management Informatics 1, 57-66 (1972).
7. Lavenberg, S.S. and C.H. Sauer, "Sequential Stopping Rules for the Regenerative Method of Simulation," IBM J. Res. Develop. 21, 545-558 (1977).
8. Law, A.M., "Confidence Intervals in Discrete Event Simulation: A Comparison of Replication and Batch Means," Naval Res. Logist. Quart. 24, 667-678 (1977).
9. Law, A.M., "Confidence Intervals for Steady-State Simulations, I: A Survey of Fixed Sample Size Procedures," Technical Report 78-5, Dept. of Industrial Engineering, University of Wisconsin (1978).
10. Law, A.M. and W.D. Kelton, "Confidence Intervals for Steady-State Simulations, II: A Survey of Sequential Procedures," Technical Report 78-6, Dept. of Industrial Engineering, University of Wisconsin (1978).
11. Nadas, A., "An Extension of a Theorem of Chow and Robbins on Sequential Confidence Intervals for the Mean," Ann. Math. Statist. 40, 667-671 (1969).

12. Starr, N., "The Performance of a Sequential Procedure for the Fixed-Width Interval Estimation of the Mean," Ann. Math. Statist. 37, 36-50 (1966).
13. Thomas, M.A., "A Simple Sequential Procedure for Sampling Termination in Simulation Investigations," J. Statist. Comput. Simul. 3, 161-164 (1974).

Table 1. Fixed Sample Size Results for $d(25|N(0)=0) = 2.12$, M/M/1 Queue with $\rho = .9$.

n	coverage	average of $\frac{\text{c.i. half length}}{\bar{X}(n)}$
5	.880 \pm .024	.672
10	.864 \pm .025	.436
20	.886 \pm .023	.301
40	.914 \pm .021	.212

Table 2. Fixed Sample Size Results for $e(12|I_1=S) = 99.52$, (s,S) Inventory System.

n	coverage	average of $\frac{\text{c.i. half length}}{\bar{X}(n)}$
5	.908 \pm .021	.048
10	.904 \pm .022	.031
20	.880 \pm .024	.021
40	.894 \pm .023	.014

Table 3. Fixed Sample Size Results for $E(T|\text{all components new}) = .778$, Reliability Model.

n	coverage	average of $\frac{\text{c.i. half length}}{\bar{X}(n)}$
5	.708 \pm .033	1.163
10	.750 \pm .032	.820
20	.800 \pm .029	.600
40	.840 \pm .027	.444

Table 4. Skewness and Kurtosis for the Three Stochastic Models and the Normal Distribution.

Stochastic Model or Distribution	Skewness	Kurtosis
Normal Distribution	0*	3*
M/M/1 Queue	1.66	6.43
(s,S) Inventory System	.45	3.76
Reliability Model	5.18	54.39

*Theoretical Values

Table 5. Relative Width Results for $d(25|N(0)=0) = 2.12$, M/M/1 Queue with $\rho = .9$.

γ	Procedure 1			Procedure 2		
	$E\{N_{r,1}(\gamma, \alpha)\}$	coverage	average c.i. half length	$E\{N_{r,2}(\gamma, \alpha)\}$	coverage	average c.i. half length
.2	42.3±0.9	.842±.027	.414	41.9±0.8	.862±.025	.437
.1	175.2±1.7	.860±.026	.211	174.5±1.7	.868±.025	.213
.05	704.4±3.5	.884±.024	.106	703.7±3.5	.882±.024	.106

Table 6. Relative Width Results for $e(12|I_1=S) = 99.52$, (s,S) Inventory System.

γ	Procedure 1			Procedure 2		
	$E\{N_{r,1}(\gamma, \alpha)\}$	coverage	average c.i. half length	$E\{N_{r,2}(\gamma, \alpha)\}$	coverage	average c.i. half length
.2	5.0±0.0	.902±.022	4.89	5.0±0.0	1.0	20.74
.1	5.0±0.0	.902±.022	4.86	5.0±0.0	1.0	10.06
.05	5.9±0.1	.892±.023	3.97	5.7±0.1	.962±.014	4.99
.025	13.3±0.4	.834±.027	2.35	12.3±0.4	.858±.026	2.48
.0125	51.0±1.0	.856±.026	1.23	49.8±1.0	.862±.025	1.24
.00625	206.3±1.8	.872±.025	0.62	205.4±1.8	.876±.024	0.62

Table 7. Relative Width Results for $E(T|\text{all components new}) = .778$, Reliability Model.

γ	Procedure 1			Procedure 2		
	$E\{N_{r,1}(\gamma, \alpha)\}$	coverage	average c.i. half length	$E\{N_{r,2}(\gamma, \alpha)\}$	coverage	average c.i. half length
.2	213.7±4.5	.876±.024	.152	214.1±4.5	.908±.021	.160
.1	907.4±11.2	.898±.022	.077	908.6±10.8	.902±.022	.078
.05	3720.5±23.7	.882±.024	.039	3720.0±23.7	.884±.024	.039

Table 8. Absolute Width Results for $d(25|N(0)=0) = 2.12$, M/M/1 Queue with $\rho = .9$.

c	Chow and Robbins		n_0	Dudewicz	
	$E\{N_{a,1}(c,\alpha)\}$	coverage		$E\{N_{a,2}(c,\alpha)\}$	coverage
.425	38.0 ± 1.2	$.800 \pm .029$	15	49.9 ± 2.1	$.850 \pm .026$
			30	48.2 ± 1.3	$.912 \pm .020$
			60	62.1 ± 0.4	$.926 \pm .019$
.212	173.5 ± 2.5	$.898 \pm .022$	15	196.9 ± 8.5	$.854 \pm .026$
			30	185.7 ± 5.6	$.888 \pm .023$
			60	182.9 ± 4.0	$.894 \pm .023$
.106	706.8 ± 4.8	$.906 \pm .021$	15	786.1 ± 34.2	$.868 \pm .025$
			30	741.1 ± 22.6	$.878 \pm .024$
			60	730.2 ± 15.7	$.898 \pm .022$

Table 9. Absolute Width Results for $e(12|I_1=S) = 99.52$, (s,S) Inventory System.

c	Chow and Robbins		n_0	Dudewicz	
	$E\{N_{a,1}(c,\alpha)\}$	coverage		$E\{N_{a,2}(c,\alpha)\}$	coverage
19.90	5.0 ± 0.0	1.0	15	16.0 ± 0.0	$.936 \pm .018$
			30	31.0 ± 0.0	$.878 \pm .024$
			60	61.0 ± 0.0	$.888 \pm .023$
9.95	5.0 ± 0.0	1.0	15	16.0 ± 0.0	$.936 \pm .018$
			30	31.0 ± 0.0	$.880 \pm .024$
			60	61.0 ± 0.0	$.890 \pm .023$
4.98	5.7 ± 0.1	$.976 \pm .011$	15	16.0 ± 0.0	$.922 \pm .020$
			30	31.0 ± 0.0	$.882 \pm .024$
			60	61.0 ± 0.0	$.886 \pm .023$
2.49	12.3 ± 0.4	$.880 \pm .024$	15	18.5 ± 0.3	$.908 \pm .021$
			30	31.0 ± 0.0	$.894 \pm .023$
			60	61.0 ± 0.0	$.882 \pm .024$
1.24	48.3 ± 1.1	$.872 \pm .025$	15	60.8 ± 2.0	$.904 \pm .022$
			30	55.0 ± 1.3	$.912 \pm .020$
			60	62.9 ± 0.4	$.902 \pm .022$
0.62	204.4 ± 1.8	$.896 \pm .022$	15	241.7 ± 8.1	$.912 \pm .020$
			30	217.9 ± 5.1	$.898 \pm .022$
			60	211.5 ± 3.4	$.912 \pm .020$

Table 10. Absolute Width Results for $E(T|\text{all components new}) = .778$, Reliability Model.

c	Chow and Robbins		n_0	Dudewicz	
	$E\{N_{\alpha,1}(c,\alpha)\}$	coverage		$E\{N_{\alpha,2}(c,\alpha)\}$	coverage
.156	179.5 \pm 7.0	.774 \pm .031	15	246.0 \pm 27.2	.704 \pm .034
			30	220.8 \pm 17.0	.772 \pm .031
			60	231.7 \pm 14.9	.812 \pm .029
.078	888.0 \pm 14.5	.900 \pm .022	15	981.8 \pm 109.0	.728 \pm .033
			30	880.6 \pm 68.0	.794 \pm .030
			60	922.2 \pm 59.6	.838 \pm .027
.039	3672.1 \pm 32.9	.884 \pm .024	15	3925.6 \pm 435.8	.772 \pm .031
			30	3520.9 \pm 272.0	.788 \pm .030
			60	3687.2 \pm 238.5	.832 \pm .028

Figure 1. $d(m|N(0)=0)$ as a Function of m for the M/M/1 Queue
with $\rho = 0.9$.

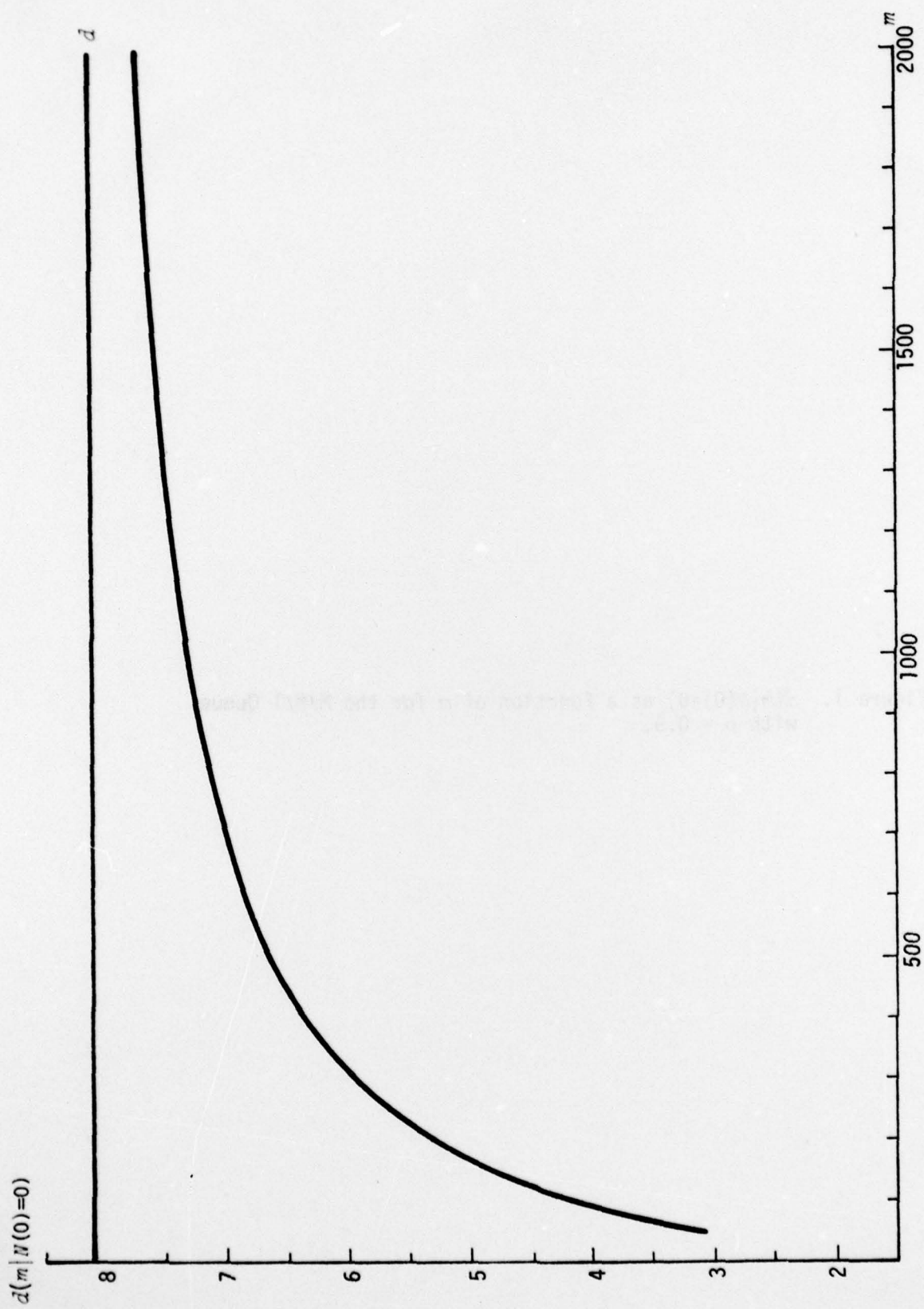


Figure 2. $e(m|I_1=s)$ as a Function of m for the (s,S) Inventory System.

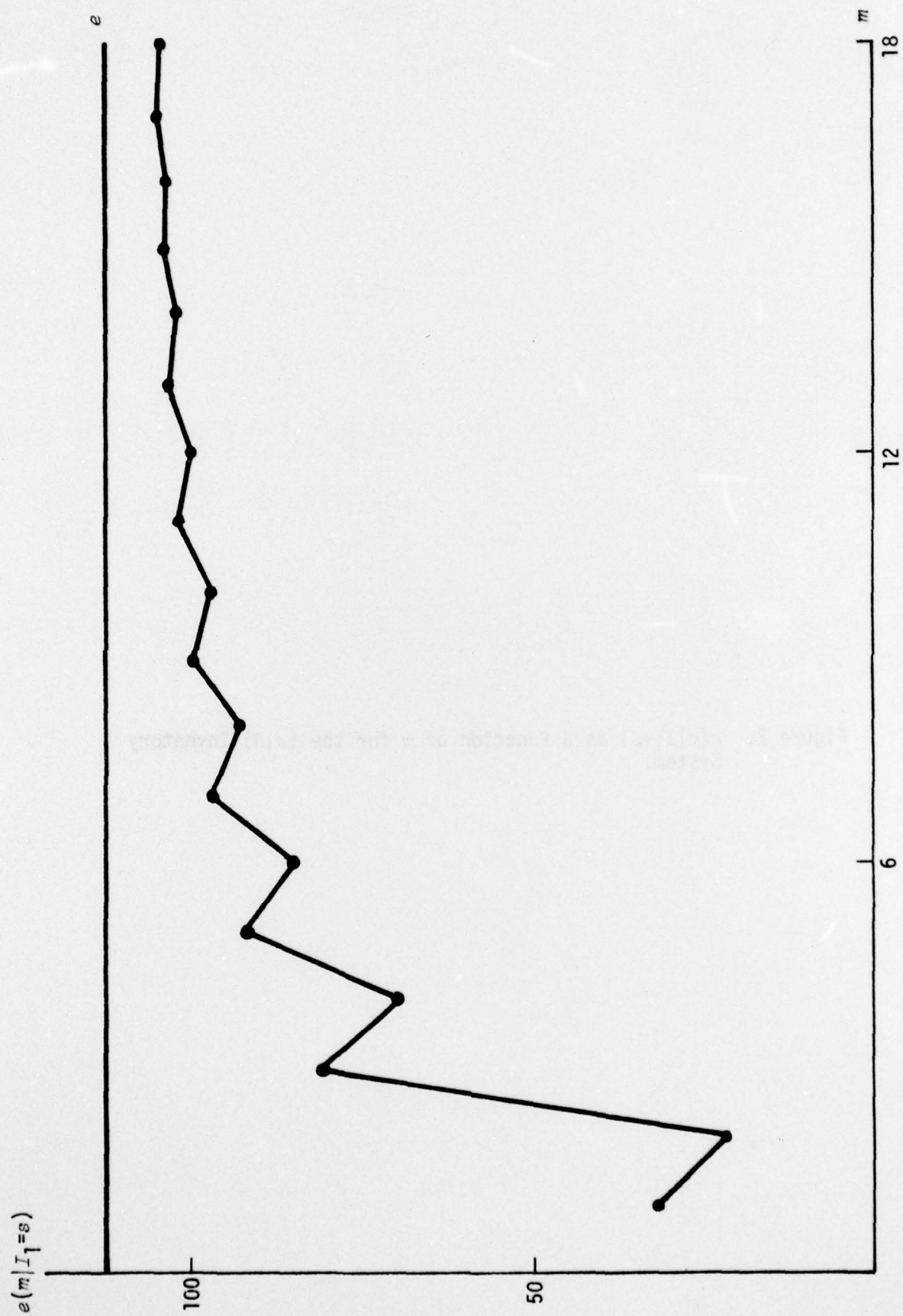
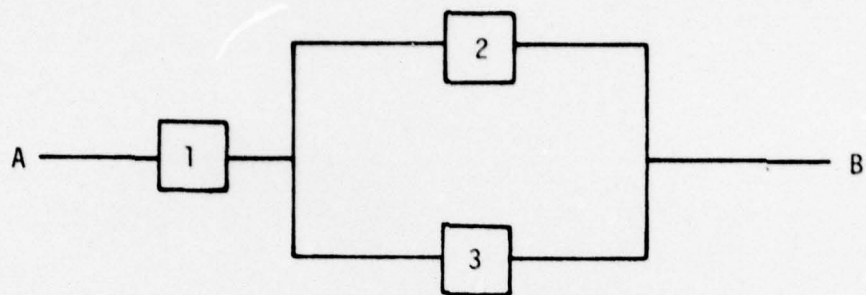




Figure 3. Reliability Model.



REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 78-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Statistical Analysis of the Output Data from Terminating Simulations		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Averill M. Law		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0403
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Wisconsin 750 University Avenue Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 047-145
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE April 1978
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Branch Office 536 S. Clark Street, Room 286 Chicago, Illinois 60605		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Simulation Confidence Interval Relative Width Procedure Absolute Width Procedure		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In this paper, we precisely define the two types of simulations (terminating and steady-state) with regard to analysis of simulation output and discuss some common measures of performance for each type. In addition, we also conclude from talking to a large number of simulation practitioners, that a significant proportion of simulations in the real world are of the terminating type. This is contrary to the impression one gets from reading the		

simulation literature, where the steady-state case is almost exclusively considered.

Although analyses of terminating simulations are considerably easier than are those of steady-state simulations, they have not received a careful treatment in the literature. We discuss and give empirical results for fixed sample size, relative width, and absolute width procedures which can be used for constructing confidence intervals for measures of performance in the terminating case.